Zeitschrift Korpora Deutsch als Fremdsprache

Jahrgang 2, Nummer 1 (Juli 2022), ISSN 2749-4411

DIE PLENARPROTOKOLLE DES DEUTSCHEN BUNDESTAGS AUF DISCOURSE LAB

Marcus Müller
TU Darmstadt

Abstract

Der Beitrag stellt das linguistisch aufbereitete Gesamtkorpus der Plenarprotokolle des Deutschen Bundestags vor. Es ist im Rahmen des Projektes Discourse Lab aufbereitet worden und wird über die Korpusanalyseplattform CQPweb für Forschung und Lehre eingesetzt. Das Korpus enthält alle Plenarprotokolle seit 1949 und ist mit linguistischen Basisinformationen sowie mit Angaben zu Redner*innen, Parteien und Redezeitpunkten angereichert. Außerdem ermöglicht eine Linkstruktur die sequentielle Analyse der Debatten.

Keywords: Korpus; Plenarprotokolle; Deutscher Bundestag; Discourse Lab

Abstract

This article presents the linguistically enhanced complete corpus of the plenary minutes of the German Bundestag. It was prepared within the framework of the Discourse Lab project and is used for research and teaching via the corpus analysis platform CQPweb. The corpus contains all plenary minutes since 1949 and is enriched with basic linguistic information as well as information on speakers, parties and speaking times. In addition, a link structure enables the sequential analysis of the debates.

Keywords: corpus; plenary minutes; German Bundestag; Discourse Lab

1. Das Korpus

Eine wichtige Ressource für die Analyse von politischer Sprache und Zeitdiskursen sind die offiziellen Protokolle der Plenardebatten des Deutschen Bundestages. Das an der TU Darmstadt und der Universität Heidelberg betriebene Discourse Lab[1] hostet ein linguistisch aufbereitetes und mit Metadaten angereichertes *Korpus der Plenarprotokolle*, das den Zeitraum von 1949 bis 2021, also alle Legislaturperioden von 1 bis 19 abdeckt. Die Protokolle werden vom Deutschen Bundestag zur Verfügung gestellt (https://www.bundestag.de/services/opendata) und werden seit 2016 von Discourse Lab für die digitale Linguistik aufbereitet[2]. Die Vorverarbeitung umfasst Tokenisierung, Satzsegmentierung, Lemmatisierung, Part-of-Speech-Tagging, die Auszeichnung der Parteizugehörigkeit der Redner*innen sowie die separate Markierung von Zwischenrufen. So können Redebeiträge mit und ohne Zwischenrufe oder auch Zwischenrufe gesondert durchsucht werden. Ein parlamentarischer

[1] Gegründet wurde Discourse Lab vom Autor dieses Textes gemeinsam mit Jörn Stegmeier. Beide sind, zusammen mit Michael Bender, an der TU Darmstadt für Discourse Lab verantwortlich. In Heidelberg wird Discourse Lab am Lehrstuhl von Ekkehard Felder koordiniert, operativ verantwortlich sind Katharina Jacob für das Germanistische Seminar und Bettina Fetzer für das Institut für Übersetzen und Dolmetschen. Daniel Wachter ist an beiden Standorten Koordinator der Korpusinfrastruktur von Discourse Lab. Zwischen 2015 und 2017 wurde Discourse Lab in der DFG-Exzellenzinitiative II, Innovationsfond FRONTIER, an der Universität Heidelberg gefördert. Für das *Korpus der Plenarprotokolle* ist insbesondere Maxine Schilde verantwortlich. Erreichbar ist Discourse Lab unter https://www.discourselab.de.

[2] In der momentanen Version sind alle Plenarprotokolle bis einschließlich 05.2022 enthalten. Das sind 809.152 Texte (Debattenbeiträge) und 261.823.225 Tokens. Das Korpus wird in regelmäßigen Abständen mit aktuellen Daten erweitert.

Redebeitrag (<text>) bildet die Grundeinheit des Korpus. Innerhalb eines Beitrags wird zwischen Text des Sprechenden (<sp>) und Zwischenruf (<z>) unterschieden. Jede <text>-Einheit besitzt neben einer eindeutigen ID folgende Metadaten: Sprecher*in, Fraktionszugehörigkeit, Legislaturperiode, Sitzungsnummer, Tag, Monat und Jahr. Abb. 1 gibt einen Überblick über das Datenmodell des Korpus.

```
<corpus>
    <text id="10_056_00006" speaker="Kleinert (Marburg)" group="GRÜNE" lp="10" session="056" day="23" month="02" year="1984">
        <sp>
            [...]
            <s>
                Die die ART
                Wahl    Wahl    NN
                ,   ,   $,
                die die PRELS
                Sie Sie PPER
                für für APPR
                morgen  morgen  ADV
                vorgesehen  vorsehen    VVPP
                haben   haben   VAFIN
                ,   ,   $,
                wird    werden  VAFIN
                vermutlich  vermutlich  ADV
                nicht   nicht   PTKNEG
                mehr    mehr    PIS
                als als KOKOM
                eine    eine    ART
                Farce   Farce   NN
                sein    sein    VAINF
                .   .   $.
            </s>
        </sp>
    </text>
    [...]
</corpus>
```

Abbildung 1
Datenmodell der Parlamentsprotokolle auf Discourse Lab

Wir sind bei der Aufbereitung von einer Nutzung ausgegangen, die immer wieder zwischen der Messung der Verteilung von Ausdrücken und Ausdruckskonstellationen, der interpretativen Textanalyse und der sequentiellen Analyse der Debatten wechselt und die Ergebnisse jeweils für die anderen Ebenen operationalisiert. Jede Belegstelle ist daher mit einem Metadatenblatt verknüpft, das neben den zuvor genannten Metadaten auch den vollständigen Sprecherbeitrag sowie Verlinkungen auf die Metadatenblätter des vorherigen und nachfolgenden Redebeitrags enthält. Das ermöglicht die Einordnung der Suchergebnisse in ihren inhaltlichen Kontext und ermöglicht es, die sequentielle Debattenstruktur für jede Sitzung in beide Richtungen leicht nachzuvollziehen. In Abb. 2 wurde auf die Darstellung des vollständigen Redebeitrags aus Platzgründen verzichtet.

Metadata for text *10_056_00006*	
Text identification code	10_056_00006
speaker name	Kleinert (Marburg)
speaker group	GRÜNE
lp	10
session	056
day	23
month	02
year	1984
speaker first	Präsident Dr. Barzel
speaker previous	Präsident Dr. Barzel
speaker next	Präsident Dr. Barzel

Abbildung 2
Metadatenblatt der Parlamentsprotokolle auf Discourse Lab

Die Metadaten können verwendet werden, um entweder Teilkorpora zu erstellen (z.B. alle Debattenbeiträge einer Fraktion) oder sich die Verteilung von Suchergebnissen anzeigen zu lassen (z.B. die

Verwendung des Lemmas Risiko in den verschiedenen Jahren/Legislaturperioden/Fraktionen). Alle XML-Attribute können zudem in Abfragen verwendet werden, um komplexere Restriktionen oder Abhängigkeiten zu formulieren. Zum Beispiel findet der folgende Ausdruck alle Tokens der Lemmata *Terminus, Terminologie, Fachausdruck, Fachwort, Begriff* sowie alle Komposita mit dem Grundwort *Begriff* außer *Kampfbegriff*, vorausgesetzt sie wurden von einem Mitglied der SPD-Fraktion im Jahr 1957 geäußert:

```
([lemma="Terminus|Terminologie|Fachausdruck|Fachwort"]|[lemma=".*be-
griff"%c&lemma!="Kampf.*"]) :: match.text_year="1957"&match.text_group="SPD"
```

2. Nutzung des Korpus

Das hier beschriebene Korpus wird in der IMS Corpus Workbench (vgl. Evert / Hardie 2011) verwaltet und browserbasiert über die Analyseumgebung CQPweb (vgl. Hardie 2012) zur Verfügung gestellt. Es kann von allen Interessierten in Forschung und Lehre eingesetzt und nach einmaliger Registrierung über eine Benutzeroberfläche durchsucht und sprachstatistisch analysiert werden. CQPweb hat vor allem einen Vorteil: Es lässt sich einerseits sehr niederschwellig einsetzen und ermöglicht aus dem Stand einfache Suchen nach Belegen und Distributionsanalysen. Andererseits lassen sich aber auch komplexe Suchen mittels einer Suchsyntax (*Corpus Query Language CQL*) durchführen und gezielt Metadaten und verschiedene Annotationsebenen ansteuern. Außerdem bietet CQPweb dem oder der Forschenden die Möglichkeit, Verfahren der inferentiellen Statistik wie Kollokationsanalysen und Keywording anzuwenden, dabei zwischen verschiedenen statistischen Maßen zu wählen und so diese Operationen auf Forschungsfragen und Korpusspezifika abzustimmen.

Als Nutzungsbeispiel sei das Verhältnis des generischen Maskulinums *Bürger* zur Doppelform *Bürgerinnen und Bürger* bei der generischen Thematisierung der entsprechenden Gruppe über die Zeit (Abb. 3 links) und in der Legislaturperiode 19 nach Parteien (Abb. 3 rechts) dargestellt. Den Graphiken liegen folgende Anfragen zugrunde[3]:

```
[word!="Bürgerin(nen)?"]{5}[word="Bürger(n|s)?"][word!="Bürgerin(nen)?"]{5}
```

Der Suchausdruck für die Doppelform ist der folgende:

```
(([word="Bürgerinnen"][]{,3}[word="und"][]{,3}[word="Bürgern?"])|([word="Bür-
gern?"] []{,3}[word="und"][]{,3}[word="Bürgerinnen"])) within s
```

[3] Die erste Anfrage bezieht auch den Singular „der Bürger“ plus Flexionsformen ein, der im Bundestag bis auf Einzelbelege immer generisch verwendet wird.

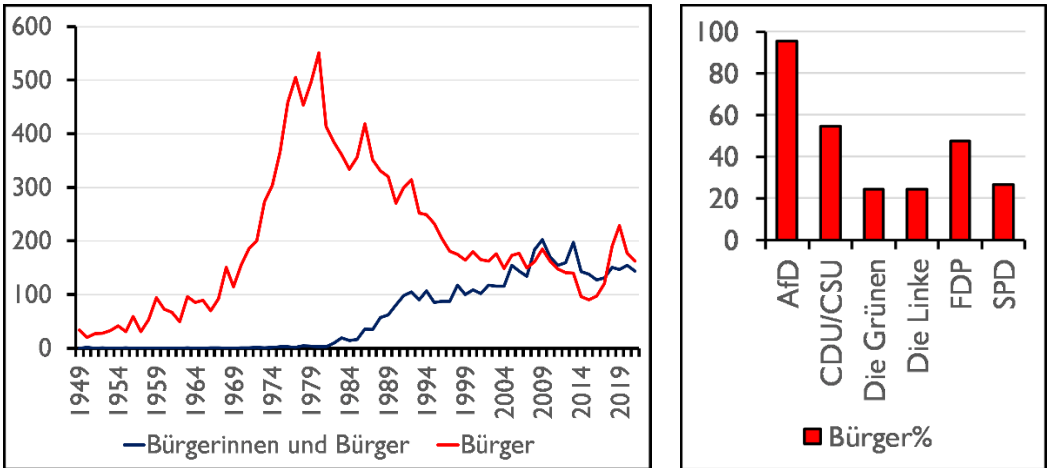

Abbildung 3
Anteil des generischen Maskulinums „Bürger“ an generischen Thematisierungen
nach Jahren (links, Frequenz je 1 Mio. Wörter)
und nach Partei in Legislaturperiode 19 (rechts, in Prozent)

Neben den wenig überraschenden Befunden, dass die Doppelform seit Mitte der 80er Jahre stetig zunimmt und dass der Anteil des generischen Maskulinums bei Vertreter*innen in der LP 19 umso größer ist, je konservativer die Partei ist, zeigt sich hier, dass ab 2018 das generische Maskulinum wieder überwiegt. Dass das vor allem an der AfD liegt, zeigt die Graphik Abb. 3 rechts. Wir sehen an ihr aber auch, dass durchaus auch Grüne und Linke in einem Fünftel der Fälle in der Legislaturperiode 2017 bis 2021von *Bürgern* sprechen, nämlich bevorzugt dann, wenn *Bürger* nicht-fokal in komplexen Nominalphrasen steht und der Redner ein Mann ist, wie im folgenden Beleg:

> *Beim Militär dürft ihr dabei sein , aber bei einem gemeinsamen ökonomischen und politischen Europa bleibt ihr* ***Bürger zweiter Klasse****.*
>
> ID 19_104_00101 – Jürgen Trittin, Die Grünen, 06.06.2019

Daneben fällt die Hausse des Konzepts BÜRGER in den 1970er Jahren auf, die Anfang der 1980er auf noch hohem Niveau beendet scheint, während der Niedergang des generischen Maskulinums *Bürger* danach zunehmend von der Doppelform aufgefangen wird – bis zum Einzug der AfD in den Bundestag.

Ausgearbeitete Nutzungsbeispiele sind neben der Studie von Müller zur Terminologiearbeit im deutschen Bundestag in diesem Heft z.B. die Studien von Müller und Mell (2021) zur Geschichte des Risikokonzeptes im politischen Diskurs und von Felder und Müller (2022) zu Moralisierungspraktiken in Bundestagsdebatten.

Das Korpus und eine E-Mailadresse zur Anmeldung sind zu finden unter: https://www.discourselab.de/cqpweb/.

Literatur und Ressourcen

Felder, Ekkehard / Müller, Marcus (2022): Diskurs korpuspragmatisch: Annotation, Kollaboration, Deutung am Beispiel von Praktiken des Moralisierens. In: Heidrun Kämper / Plewnia, Albrecht (Hrsg.): *Sprache in Politik und Gesellschaft. Perspektiven und Zugänge* (IDS Jahrbuch 2021). Berlin/Boston: De Gruyter, 241–261.

Evert, Stephan / Hardie, Andrew (2011): Twenty-first Century Corpus Workbench: Updating a Query Architecture for the New Millenium. In: *Proceedings of the Corpus Linguistics 2011 Conference, University of Birmingham*. http://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-153.pdf (24.02.2022).

Hardie, Andrew (2012): CQPweb — Combining Power, Flexibility and Usability in a Corpus Analysis Tool. In: *International Journal of Corpus Linguistics* 17: 3, 380-409.

Müller, Marcus / Mell, Ruth M. (2021): 'Risk' in Political Discourse. A Corpus Approach to Semantic Change in German Bundestag Debates. In: *International Journal of Risk Research*. DOI: 10.1080/13669877.2021.1913631.

Müller, Marcus / Stegmeier, Jörn (Hrsg.) (2018): *Korpus der Plenarprotokolle des deutschen Bundestags*.

Legislaturperiode 1-20. CQPweb-Edition. Discourse Lab. Darmstadt. https://www.discourselab.de/CQPweb/ (24.02.2022).

Biografische Notiz: Marcus Müller ist Professor für Germanistik – Digitale Linguistik an der Technischen Universität Darmstadt. Zu seinen Forschungsschwerpunkten gehören digitale Diskursanalyse, Korpuslinguistik, Wissenschaftsdiskurse, grammatische Variation sowie Sprache und Kunst. Aktuelle Arbeitsschwerpunkte sind empirische Terminologieforschung, öffentliche Risikodiskurse in Deutschland und Großbritannien und heuristische Praktiken in wissenschaftlichen Texten.

Kontaktanschrift:
Prof. Dr. Marcus Müller
TU Darmstadt
Institut für Sprach- und Literaturwissenschaft
Residenzschloss, Marktplatz 15
64283 Darmstadt
marcus.mueller@tu-darmstadt.de